

Aplicabilidad de EAD en archivos históricos de ingeniería: el fondo Carlos Fernández Casado

Applying EAD to historical engineering archives: the holdings of Carlos Fernández Casado

Ricardo EITO BRUN

Universidad Carlos III de Madrid, c/ Madrid 124, Getafe (Madrid), reito@bib.uc3m.es

Resumen

En este artículo se describe el desarrollo de un entorno de trabajo basado en estándares XML, concretamente EAD y EAC-CPF, para la gestión y publicación de las descripciones y documentos del fondo del Ingeniero Carlos Fernández Casado. El trabajo ha sido realizado por un equipo de investigadores de la Universidad Carlos III de Madrid, Ricardo Eito Brun y Antonio Perianes, para el Centro de Estudios Históricos de Obras Públicas y Urbanismo (CEHOPU) del Ministerio de Fomento de España. Como parte de la solución propuesta, se ha desarrollado un entorno de edición que permite crear instrumentos de descripción EAD compatibles con ISAD(G) y registros de autoridad y de contexto EAC-CPF compatibles con ISAAR(CPF). El entorno de trabajo incorpora funciones adicionales como el acceso a vocabularios controlados locales y remotos para la asignación de descriptores, la capacidad de vincular objetos digitales a las descripciones y utilidades para la publicación web de las descripciones EAD.

Palabras clave: EAD. EAC-CPF. Archivos de ingeniería. Thesauri. ISAD(G). ISAAR(CPF)

Abstract

This paper describes the development of a collaborative work environment based on XML standards (EAD and EAC-CPF) for the management and publication of finding aids and documents from the personal holdings of the Spanish engineer Carlos Fernández Casado. This project has been completed by a research team from Universidad Carlos III de Madrid led by Ricardo Eito Brun and Antonio Perianes for Centro de Estudios Históricos de Obras Públicas y Urbanismo (CEHOPU), an institution that depends on the Ministerio de Fomento de España. As part of the technical solution, the research team has developed an editing tool for the creation of EAD finding aids compatible with ISAD(G) and EAC-CPF authority and context records compatible with ISAAR(CPF). The work environment also incorporates additional functions to access local and remote controlled vocabularies and thesauri, as well as the capability of linking digital objects to the finding aids and publishing them through the Web.

Keywords: EAD. EAC-CPF. Engineering archives. Thesauri. ISAD(G). ISAAR(CPF).

1. El fondo

El fondo personal del ingeniero Carlos Fernández Casado se incorporó al Centro de Estudios Históricos de Obras Públicas y Urbanismo (CEHOPU) en 2008. Los documentos proceden de la oficina creada por Fernández Casado en 1966 y de la familia del ingeniero (Fernández Troyano, 2007), y abarca materiales representativos de su actividad profesional y docente. En 2006 CEHOPU acometió una primera fase de inventario y clasificación de los proyectos y obras, y tras la cesión de 2008 se impulsó la catalogación y digitalización sistemática de los fondos. Los trabajos comenzaron por la documentación procedente de la Oficina Técnica y, al poco tiempo y en paralelo, se puso en marcha el proyecto de Diseño y desarrollo del entorno de gestión del Archivo, desarrollado por un equipo de investigadores del Departamento de Biblioteconomía y Documentación de la Universidad Carlos III de Madrid, que permite en la actualidad consultar el material catalogado a

través de la web de CEHOPU. El fondo consta principalmente de expedientes de obras, planos y fotografías. CEHOPU facilitó al equipo investigador bases de datos Microsoft® Access® con las descripciones de obras e imágenes, y una colección de documentos Word® con la descripción de las distintas unidades documentales (expedientes de obra, planos, etc.)

La primera actividad del proyecto fue analizar la aplicabilidad del modelo EAD para describir los materiales del fondo. Se estableció una correspondencia entre los campos de la base de datos Access y los elementos del modelo EAD (compatible con ISAD(G)). También se estableció la correspondencia entre la información disponible en los documentos Word para las unidades documentales y los elementos EAD.

2. Aproximación técnica

Los objetivos establecidos por CEHOPU con esta colaboración se resumen en los siguientes

puntos: 1) disponer de un entorno de trabajo que facilitase la creación de descripciones acordes con ISAD(G); 2) asegurar la reutilización de descripciones creadas con motivo de una exposición sobre la actividad del ingeniero desarrollada con anterioridad; 3) disponer de un entorno que permitiese la publicación progresiva de las descripciones y que se pueda usar para otros fondos; y 4) poder reutilizar las descripciones si se decide utilizar otra herramienta software en el futuro.

La aproximación técnica escogida se basó en el uso de formatos estándares, abiertos EAD y EAC-CPF, basados en XML.

2.1. Aplicación de EAD en archivos históricos de ingeniería

En líneas generales, la aproximación seguida ha sido la siguiente:

1. Para cada obra, se ha creado un documento EAD independiente, donde se recoge la descripción de las distintas obras en las que participó el ingeniero.
2. Para cada agrupación de documentos (expedientes de obra), se ha creado un documento EAD independiente. La información disponible en estos documentos se corresponde con las descripciones facilitadas por el CEHOPU en los documentos Word antes mencionados.
3. Para cada imagen digitalizada (fotografía, plano, etc.) se ha creado un documento EAD independiente con la descripción y las características de cada imagen individual. La aplicabilidad de EAD para describir colecciones de imagen ha sido analizada con anterioridad en la bibliografía profesional (Elings, 1998; Higgins, 1998).

El modelo EAD permite la descripción multinivel y establecer relaciones entre los registros EAD que corresponden a las distintas unidades de descripción. Concretamente, se han establecido estas relaciones:

1. Se ha considerado que el registro de cada obra corresponde a una colección, donde se agrupan unidades documentales de niveles inferiores correspondientes a las imágenes digitalizadas. En EAD, esto se ha materializado añadiendo a los documentos de cada obra un elemento <dsc> donde se han añadido elementos <c> (uno para cada imagen). Los documentos EAD correspondientes a la obra se han catalogado como colecciones, ya que no se corresponden con agrupaciones

orgánicas de documentos, según el concepto archivístico.

2. En los registros correspondientes a cada obra se ha añadido una referencia a las unidades documentales (expedientes de obra), mediante el elemento EAD <archref>, que permite establecer relaciones con otras agrupaciones o unidades documentales disponibles. De esta forma, en el documento EAD correspondiente a una obra podemos encontrar referencias o hiperenlaces a las unidades documentales y a las imágenes relacionadas con dicha obra.

Este procedimiento ha resultado adecuado y ha sido acordado con el personal responsable de la gestión del archivo del CEHOPU, ya que desde la perspectiva del usuario el principal punto de acceso para la información es la obra. En una aproximación inicial, el equipo investigador pensó que era más acorde con la teoría archivística tomar las series y los expedientes como punto de acceso principal, pero esta aproximación no corresponde al comportamiento que siguen los usuarios. También existía el problema de la vinculación de los elementos gráficos a los expedientes (éstos estaban vinculados a las obras, pero no a los expedientes). Por este motivo, se optó por el planteamiento descrito en los párrafos anteriores.

Para asegurar que los documentos EAD son autocontenidos y evitar cualquier duplicación de información innecesaria, los EAD creados para los expedientes de obra se vincularon a las obras mediante el elemento <name>.

En el caso de los documentos EAD correspondientes a imágenes digitalizadas, éstos recogen la relación entre la imagen y la obra mediante un elemento <name> dentro de <controlaccess>. No existe una vinculación entre las imágenes y los expedientes de obra.

A parte de esto, en todos los casos se mantiene la relación entre obras, expedientes e imágenes, a partir de las convenciones seguidas en los nombres de los ficheros. En éstos se ha seguido la codificación propuesta por CEHOPU, de forma que siempre es posible identificar qué elementos o registros están relacionados entre sí y cuáles están relacionados con una misma obra. En los EAD se ha utilizado el elemento <unitid> para recoger el código usado para nombrar a los ficheros.

Las características físicas de la imagen se han codificado en el elemento <physdesc>. Concretamente, dentro de este elemento se ha usado el elemento <physfacet> para los distintos aspectos que se recogían en la base de datos

Access. El dato que se recoge se distingue mediante el atributo @type, donde se indica el tipo de información.

El elemento <controlaccess> recoge los puntos de acceso o descriptores usados en la base de datos, distinguiendo mediante el atributo @role de qué tipo de dato se trata (concretamente, en el caso de descriptores correspondientes a nombres geográficos, se indica sí se trata de provincias, comunidades autónomas, etc.

Para vincular los registros EAD correspondientes a las imágenes con los archivos de imagen correspondientes se han añadido elementos <daogrp>, que contienen un elemento <extrefloc> para cada archivo de imagen. Se ha distinguido entre archivos de imagen de baja resolución e imágenes en miniatura (thumbnails).

La figura 1 corresponde a uno de los documentos generados para una de imagen. La descripción EAD se ha generado a partir de un proceso de conversión —basado en hojas de estilo XSLT— desde los registros de la base de datos Access de CEHOPU. Para facilitar la conversión de Access a XML se ha usado Altova® XML Spy®. Esta herramienta permite crear proyectos donde se agrupan los documentos XML a convertir, las hojas de estilo XSLT y ejecutar la conversión de un amplio conjunto de archivos XML de forma desatendida. Estos programas se han usado durante la fase de diseño y para el proceso de conversión, pero no son necesarios para el mantenimiento y gestión posterior de la publicación.

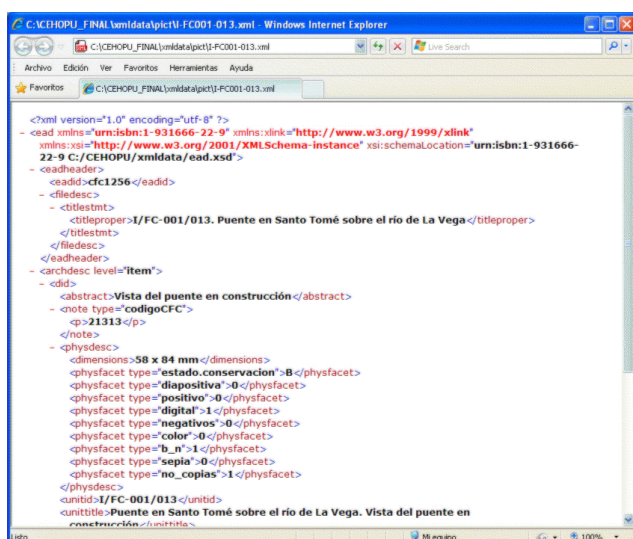


Figura 1. Ficha de imagen codificada en EAD

La figura 2 corresponde a uno de los documentos EAD para una obra. La descripción de las

obras se ha generado a partir de un proceso de conversión —basado en hojas de estilo XSLT— desde los registros de la base de datos Access de CEHOPU.

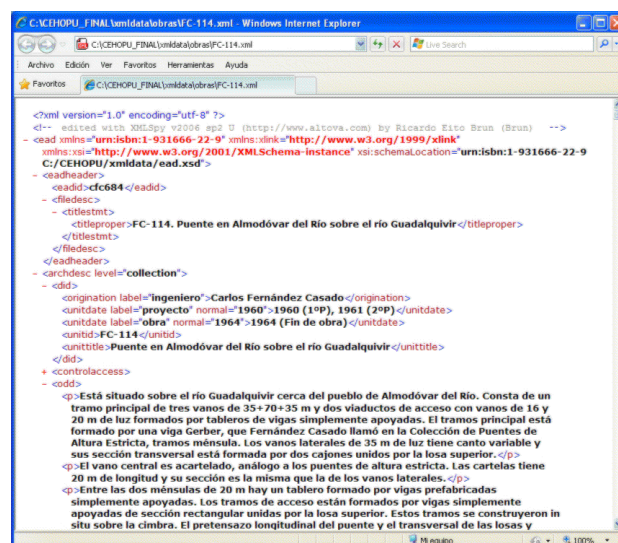


Figura 2. Ficha de obra codificada en EAD

En este caso, señalaremos como rasgos más característicos el uso de elementos <odd> para codificar la descripción de las obras. Tras el análisis inicial de la correspondencia entre las descripciones existentes y EAD se optó por usar el elemento *Other Descriptive Data* para recoger esta información. Junto a esto, al considerar que los documentos EAD de las obras corresponden a colecciones (esto es, a una agrupación artificial de documentos), en estos EAD se ha añadido un elemento <dsc> para agrupar las referencias a las imágenes relacionadas, con un componente <c> independiente para cada imagen. Con el objeto de facilitar la generación de las páginas de visualización HTML se ha desarrollado una utilidad software que realiza una fusión entre la descripción de cada obra y las imágenes correspondientes a la misma. Un proceso similar se aplica para vincular los EAD de las obras con los EAD de sus expedientes relacionados.

El concepto de expediente (unidad documental compuesta) corresponde a la descripción de grupos de documentos relacionados con una obra facilitados por CEHOPU al equipo investigador mediante documentos Word. En este caso, se ha aplicado un proceso de conversión de los documentos Word semiautomatizado. Para los expedientes se ha optado por recoger el contenido de mediante una descripción textual dentro del elemento <scopecontent>. Un punto destacable es la implementación de un

proceso automatizado que —a partir de las convenciones de nombres seguidas para nombrar a estos archivos— asocia mediante elementos <archref> los expedientes con las obras.

Estas referencias cruzadas entre expedientes relacionados, expedientes y obras, obras e imágenes, etc., codificadas en los documentos EAD, se traducen posteriormente en enlaces hipertexto HTML que permiten recorrer y navegar la publicación.



Figura 3. Enlaces hipertexto entre obras, expedientes y descripciones de imagen

2.2. Gestión de índices

A partir de los documentos XML de obras, imágenes y expedientes, se han generado mediante un proceso automatizado varios índices para facilitar el acceso y la recuperación. Concretamente, se han generado cuatro índices: topográfico, para lugares; instituciones y personas; tipología de obra y materias; y cronológico.

Estos índices se generan de la siguiente forma:

- El índice topográfico se genera a partir de los elementos <geogname> de los EAD correspondientes a las obras.
- El índice de instituciones y personas se genera a partir de los elementos <corpname> y <persname> de los EAD correspondientes a las obras y expedientes.
- El índice de tipología de obras y materias se genera a partir de los elementos <subject> de los EAD correspondientes a las obras.
- El índice cronológico se genera a partir del elemento <unitdate> y su atributo @normal, de los archivos EAD correspondientes a las obras.

Para cada índice se genera un documento XML (no basado en EAD, sino en un esquema propietario). La siguiente imagen muestra un ejemplo del índice topográfico (todos los índices siguen una estructura similar, salvo el índice de obras).

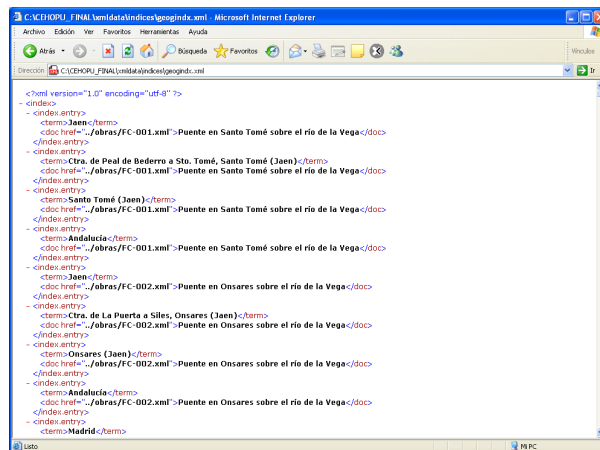


Figura 4. Estructura de los índices

En cuanto a la generación de estos índices, en primer lugar se generan los XML correspondientes extrayendo de la colección de documentos las entradas del índice. Estas no se añaden ordenadas por ningún criterio: conforme se identifican descriptores o términos de indización en los documentos, estos se añaden al índice. En una segunda etapa se reordenan las entradas del índice a partir de los elementos usados como claves de ordenación (nombres de lugares, de personas, años, etc.) Esta reordenación se realiza mediante una hoja de estilo XSLT. El resultado de aplicar estas hojas de estilo es otro archivo XML con las mismas entradas que el índice inicial, pero ordenadas alfabéticamente por su clave de ordenación.

2.3. Generación de la publicación web

El equipo de trabajo ha diseñado distintas hojas de estilo XSLT para obtener las páginas HTML que formarán la publicación web final. Las hojas de estilo XSLT se diseñaron con las herramientas Altova® StyleVision® y XML Spy®. Se ha aplicado esta misma tecnología para generar los ficheros HTML correspondientes a los índices.

2.4. Entorno de edición

La edición de los contenidos existentes y la creación de nuevos contenidos, debe hacerse directamente en formato EAD/XML. Con este fin, se han desarrollado distintas plantillas para la edición de documentos basadas en el editor

gratuito Altova Authentic®. El proceso de edición obliga a considerar no solo aspectos técnicos, sino también los relativos a la coordinación de las actividades que aseguren la coherencia en la catalogación e indización (Smith, 2008).

La plantilla tiene un aspecto como el siguiente:

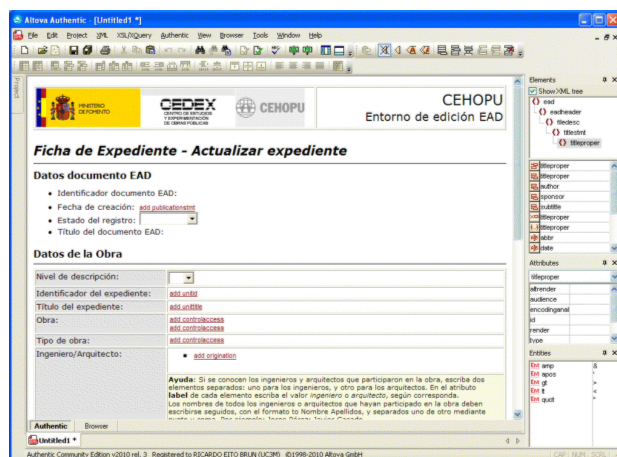


Figura 5. Herramienta para crear descripciones EAD

De esta forma, se evita que el personal de CEHOPU encargado de crear las descripciones tenga que conocer los detalles sobre el etiquetado XML y EAD, y se ofrece la posibilidad de usar un editor intuitivo para crear y modificar las descripciones. El editor ofrece la posibilidad de consultar distintos vocabularios controlados locales y remotos basados en SKOS (Miles, 2007; Pastor, 2009), y asignar descriptores procedentes de estos thesauri al instrumento de descripción que se está editando.

El alcance del proyecto ha incluido la migración a SKOS de las listas de descriptores y vocabularios controlados usados por CEHOPU. Para permitir el acceso y consulta remota a estos vocabularios controlados, se han generado bases de datos XML nativas con Oracle dbxml®, y una interfaz para el acceso basada en el proto-

colo técnico SRU (Search/Retrieve URL) publicado por la Library of Congress norteamericana.

Conclusiones

En este trabajo se resumen las principales características del proyecto de conversión a EAD y publicación en la web de las descripciones del fondo personal del ingeniero Carlos Fernández Casado. En este proyecto, se ha analizado la aplicabilidad del estándar EAD para describir obras de ingeniería, documentos y agrupaciones de documentos de ingeniería. Los resultados obtenidos son satisfactorios, y demuestran que este estándar de descripción puede utilizarse en ámbitos adicionales a la descripción de documentos de archivo.

El entorno de trabajo propuesto tiene como ventaja el uso intensivo de los estándares abiertos XML y XSLT, lo que permite reutilizar contenidos y aplicar la solución propuesta para la gestión y publicación de otros fondos de características similares al fondo objeto de este estudio.

Referencias

- Elings, M. W.; Garcelon, E. (1998). The Robert Honeyman Jr. Collection Digital Archive: EAD and the Use of Library and Museum Descriptive Standards. // *Archives and Museum Informatics*. 12, 205-219.
- Fernández Troyano, L. (2007). Carlos Fernández Casado. Ingeniero, Madrid, Ministerio de Fomento, CEHOPU.
- Higgins, R. (1998). Case Study of EAD Implementation at Durham University Library Archives and Special Collections. // *Archives and Museum Informatics*. 12, 221-234.
- Miles, A.; Perez-Agüera, J.R. (2007). SKOS: Simple Knowledge Organisation for the Web. // *Cataloging & Classification Quarterly*. 43:3/4, 69-83.
- Pastor-Sanchez, J. A.; et al. (2009). Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. // *Information Research*. 14:4, 19-19.
- Smith II, P. L. (2008). Preparing Locally Encoded Electronic Finding Aid Inventories for Union Environments: A Publishing Model for Encoded Archival Description. // *Information Technology and Libraries*. 27:2, 26-30.

